STA 380, Part 2: Exercises

Aishwarya Rajeev

2022-08-13

Introduction We chose the below ETFs to provide portfolio diversification and a range of risk.

- Invesco (QQQ), one of the biggest, exclusively non-financial stock, and heavily tech-heavy trusts.
- SPY is one of the safest and largest ETF.
- iShares Russell 1000 Growth ETF (IWF) is one of the most popular US large-cap growth ETFs with a long track record. It is an aggressive growing ETF *SCO a low performing stock

In total, we have selected 3 ETFs - "QQQ", "SPY", "IWF". We looked at data from ETFs for five years commencing on January 1, 2017

Sample Data for QQQ

```
##            QQQ.Open QQQ.High  QQQ.Low QQQ.Close QQQ.Volume QQQ.Adjusted
## 2017-01-03 114.6697 115.3619 114.3043  114.9293   22307600     114.9293
## 2017-01-04 115.0542 115.7657 115.0446  115.5542   19749100     115.5542
## 2017-01-05 115.4677 116.2849 115.4677  116.2080   20644300     116.2080
## 2017-01-06 116.3330 117.5347 116.0349  117.2271   24074300     117.2271
## 2017-01-09 117.3232 117.8232 117.2463  117.6117   18909200     117.6117
## 2017-01-10 117.6597 118.2270 117.3905  117.8712   16176600     117.8713
```

Combine all the returns in a matrix

```
##                ClCl.QQQa      ClCl.SPYa      ClCl.IWFa     ClCl.SCOa
## 2017-01-04 0.005437519  0.0059491963   7.758539e-03 -0.022081119
## 2017-01-05 0.005657717 -0.0007944567   1.596075e-03 -0.020414402
## 2017-01-06 0.008769728  0.0035777960   6.467970e-03  0.002683852
## 2017-01-09 0.003280587 -0.0033009109  -9.315451e-05  0.060620377
## 2017-01-10 0.002207112  0.0000000000   3.725876e-04  0.042903838
## 2017-01-11 0.002691697  0.0028261017   7.448883e-04 -0.053096068
```
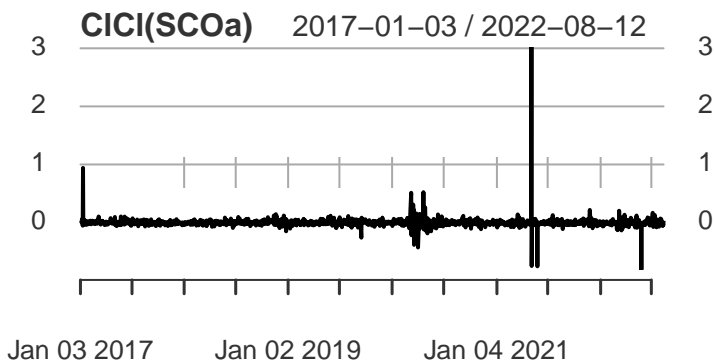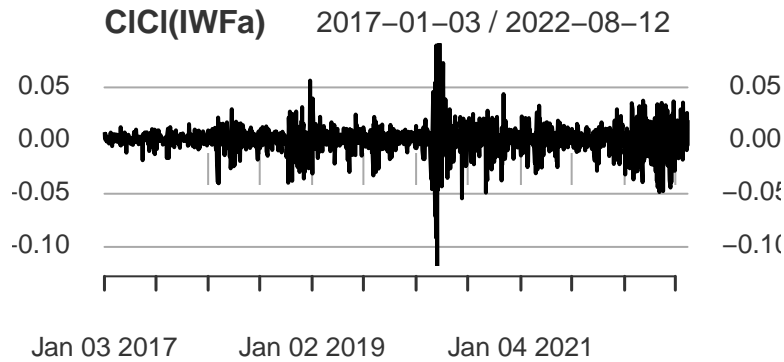
Compute the returns from the closing prices
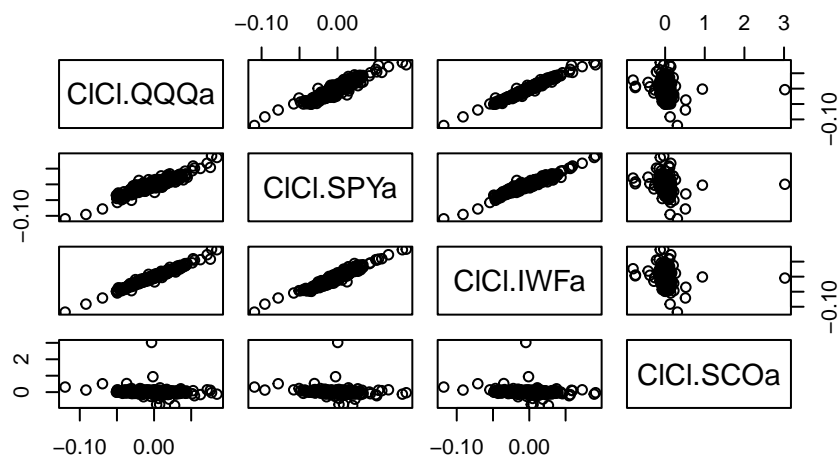
- We can see a strong linear correlation here in 3 of the stocks "QQQ", "SPY" & "IWF.""SCO" seems to not have a strong correlation with the other stocks

**Volatility of the ETFs across the 5 year period.**

CICI(IWFa)   2017–01–03 / 2022–08–12

CICI(SCOa)   2017–01–03 / 2022–08–12
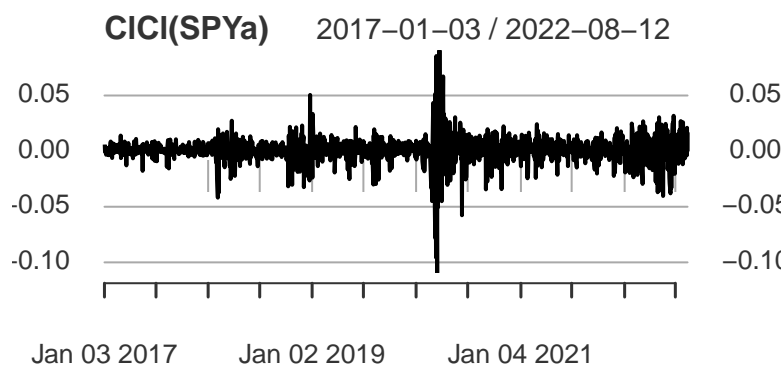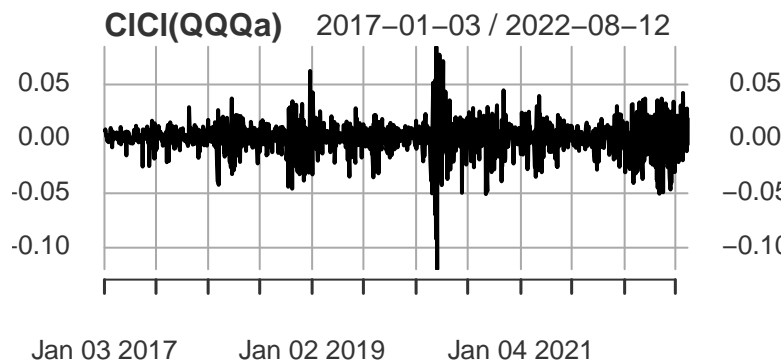
Sample a random return from the empirical joint distribution

```
##               ClCl.QQQa   ClCl.SPYa   ClCl.IWFa   ClCl.SCOa
## 2021-12-23 0.00753903 0.006222079 0.007965898 -0.02472723
```

INITIAL INVESTMENT IS $100000 SIMULATION 1: LOW RISK PORTFOLIO

Average return of investement after 20 days - $101190 5% Value at Risk for safe portfolio - $12139.7

```
##                 [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## result.1 100056.51 102552.2 102073.54 189314.84 186699.3 184761.26 181240.74
## result.2  99943.96 101260.8 102504.60 102498.98 102250.7 102715.66 104529.99
## result.3  99606.78  98882.7  99341.46  99469.41  99114.1  99738.34  99203.91
## result.4  99252.83 100371.2 100932.99 101985.95 100807.7  99659.26  99232.28
## result.5  99835.70 100234.2 174651.88 170101.12 167937.1 171708.67 175671.64
## result.6 100156.37 100199.7 102632.74 103294.58 102638.3 102507.11 102113.24
##                 [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
## result.1 171537.98 170707.05 168562.43 165184.56 166514.08 166884.91 169455.77
## result.2 103319.31 101360.89 101684.36 100857.85 101718.86 101446.15 100972.06
## result.3  99555.90  99875.20  99471.74  98377.89  95954.34  93242.68  92809.79
## result.4  98957.23  98939.82  98782.85  99140.56  99175.22 100138.46 101134.76
## result.5 168970.62 171820.72 175599.34 176342.45 175979.18 162384.43 171424.69
## result.6 101173.29  99167.88  98529.24 100018.80 100097.24  99616.58  99889.75
##                [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
## result.1 171309.84 172758.59 173170.46 171278.63 166798.45 173585.02
## result.2 102443.98 103024.41  97744.75  96393.73  97512.66  96135.83
## result.3  92675.07  92955.49  94275.13  94271.07  94865.97  95478.56
## result.4 100187.92 101841.73 101806.00 101141.75 103345.60 102222.76
## result.5 173378.46 171862.85 171573.40 171070.60 179074.15 168631.36
## result.6  99506.54 100372.01  99644.63 100969.25 100598.85 102839.00
```

Histogram of sim1[, n_days]

density.default(x = sim1[, n_days])

N = 5000 Bandwidth = 1016
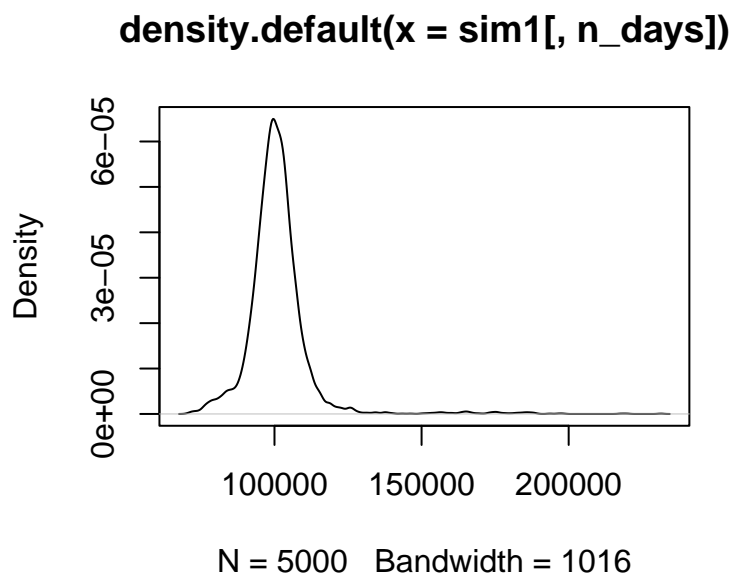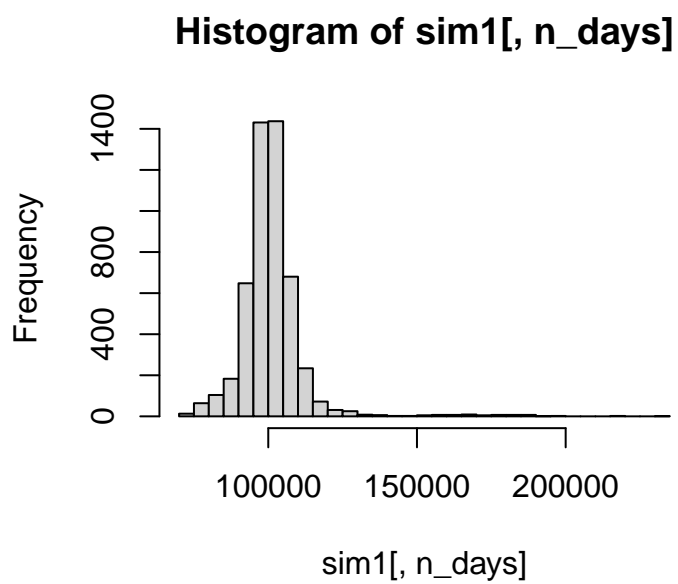
Profit and Loss

```
##
## Average return of investement after 20 days 101038.1
##
## Average profit/loss after 20 days 1038.065
```
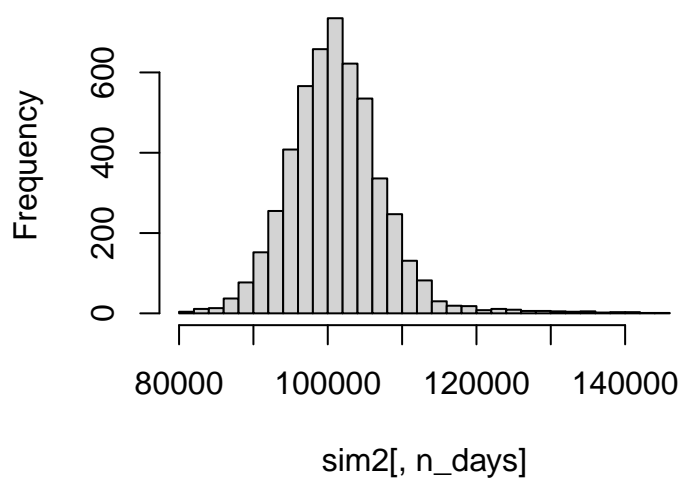
Histogram of sim1[, n_days] – initial_wealth
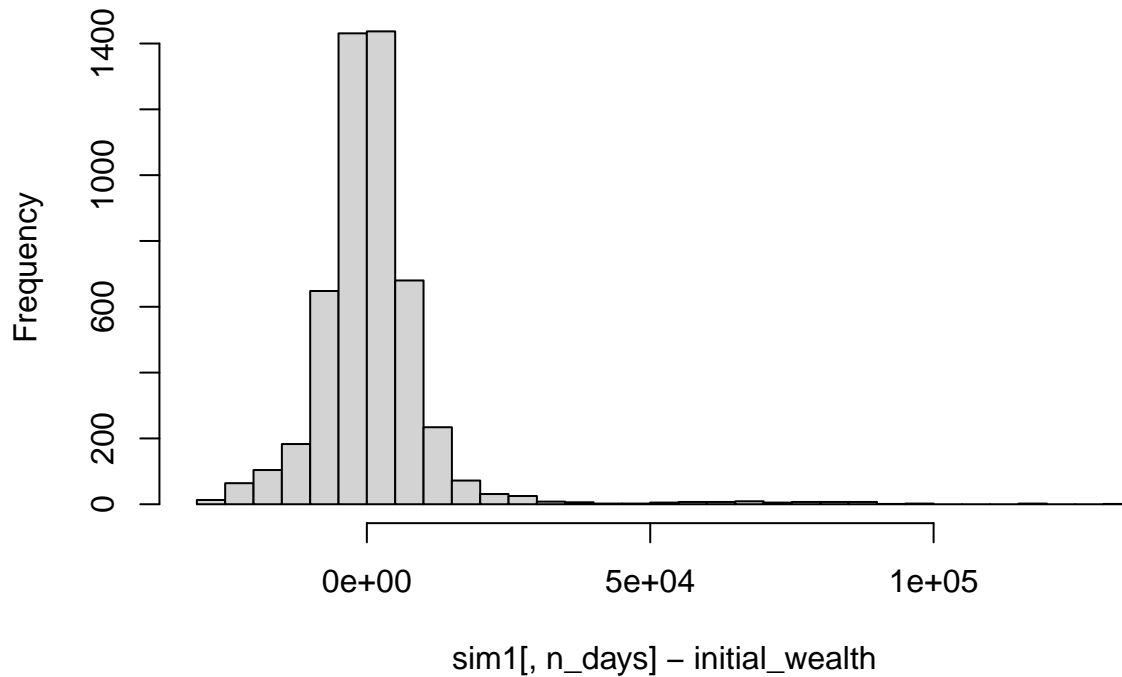

```
##
## 5% Value at Risk for the first simulation- -12465.28
```

SIMULATION 2: HIGH RISK PORTFOLIO
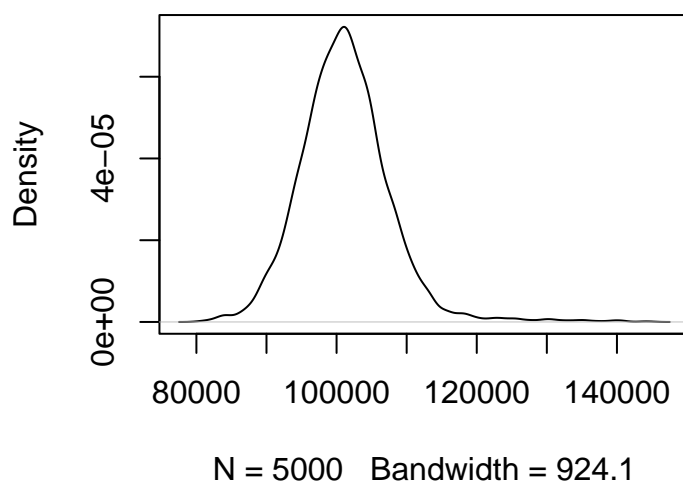
Average return of investement after 20 days - $101576.2 5% Value at Risk for safe portfolio - $7849.9327

```
##                 [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## result.1 102045.52 101712.07 100515.22 101401.98 100242.75 100696.72  99188.98
## result.2  98904.90  99020.05  94968.64  92724.08  92118.42  91996.01  92131.77
## result.3  98878.28  98273.61  97618.06  94088.97  95172.35  93094.99  92833.07
## result.4  99432.00 100743.92 101666.22 102297.94 103436.21 102525.83 102581.16
## result.5 100547.75 100907.22 101631.06  99613.14  99261.39 102462.43 105107.28
## result.6 101205.38 101611.61 100593.64 101527.70 101893.02 102381.37 101959.54
##                 [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
## result.1  99078.43  99753.48  99244.24  99554.09  99422.80  99751.42  99509.87
## result.2  92777.83  93825.58  92866.67  92314.48  92042.17  92403.70  97920.16
## result.3  93102.09  93906.66  92845.03  93223.73  92979.12  91259.86  90830.24
## result.4  99657.11 100861.77 100947.11 103756.13 103928.01 104329.23 104022.72
## result.5 105014.72 104506.47 104897.61 106332.82 106114.74 105804.98 106285.91
## result.6 101164.07 101144.72 100680.94 101007.33 100254.67 100341.24 100827.74
##                [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
## result.1  98906.02  98514.26  99040.26  98384.52  98022.67  93795.59
## result.2  97750.76  97669.65  96867.00  96830.77  96766.59  96967.60
## result.3  91045.99  91631.56  91316.70  93291.25  93168.84  93375.97
## result.4 103782.74 103042.36 104460.46 104780.76 103242.43 102720.91
## result.5 106848.22 107108.10 107410.56 106786.30 108228.80 108433.20
## result.6 100142.77  99894.72  99324.58  99462.48  98919.94  99395.86
```
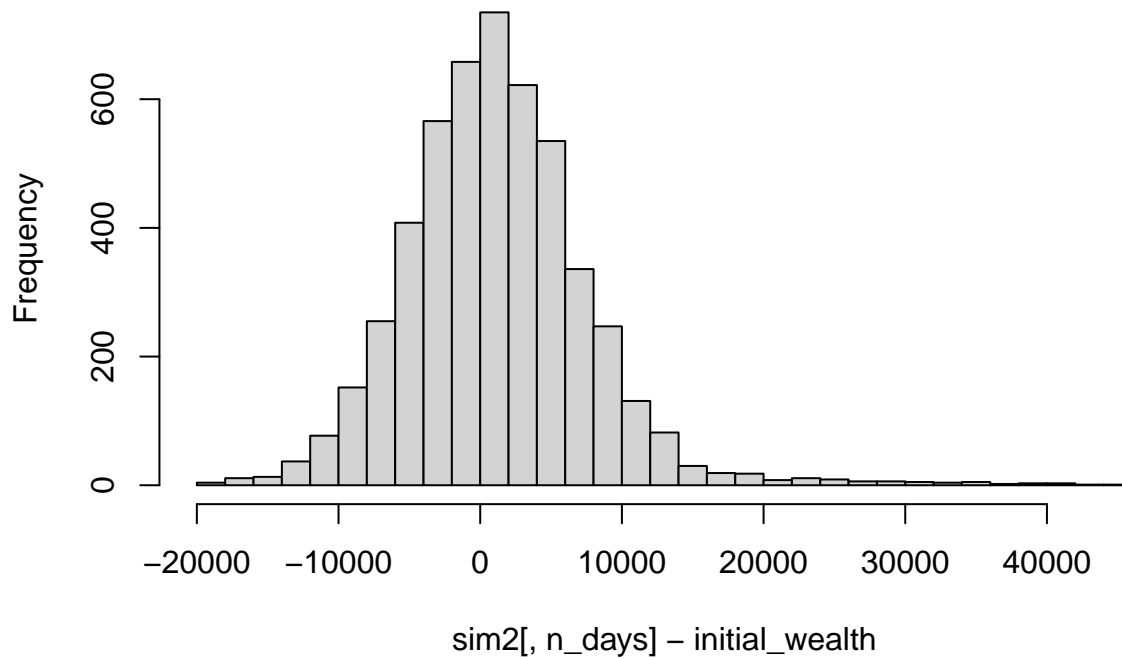
Histogram of sim2[, n_days]

density.default(x = sim2[, n_days])

N = 5000   Bandwidth = 924.1

Profit and Loss

```
##
## Average return of investement after 20 days 101201.5
##
## Average profit/loss after 20 days 1201.467
```

Histogram of sim2[, n_days] – initial_wealth

** Value at Risk **

```
##
## 5% Value at Risk for the first simulation- -8473.227
```

Summary For the safe portfolio, we are observing lower return of investment and lower 5% VaR. As the portfolio risk increased, we are able to witness the increase in returns and an increase in VaR value as expected

References: https://www.bankrate.com/investing/best-etfs/ https://etfdb.com/compare/lowest-ytd-returns/

Clustering and PCA   Sample Data

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1           7.4             0.70        0.00            1.9     0.076
## 2           7.8             0.88        0.00            2.6     0.098
## 3           7.8             0.76        0.04            2.3     0.092
## 4          11.2             0.28        0.56            1.9     0.075
## 5           7.4             0.70        0.00            1.9     0.076
## 6           7.4             0.66        0.00            1.8     0.075
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  11                   34  0.9978 3.51      0.56     9.4
## 2                  25                   67  0.9968 3.20      0.68     9.8
## 3                  15                   54  0.9970 3.26      0.65     9.8
## 4                  17                   60  0.9980 3.16      0.58     9.8
## 5                  11                   34  0.9978 3.51      0.56     9.4
## 6                  13                   40  0.9978 3.51      0.56     9.4
```

```
##   quality color
## 1       5   red
## 2       5   red
## 3       5   red
## 4       6   red
## 5       5   red
## 6       5   red
```

Data Summary

```
##  fixed.acidity    volatile.acidity  citric.acid      residual.sugar
##  Min.   : 3.800   Min.   :0.0800   Min.   :0.0000   Min.   : 0.600
##  1st Qu.: 6.400   1st Qu.:0.2300   1st Qu.:0.2500   1st Qu.: 1.800
##  Median : 7.000   Median :0.2900   Median :0.3100   Median : 3.000
##  Mean   : 7.215   Mean   :0.3397   Mean   :0.3186   Mean   : 5.443
##  3rd Qu.: 7.700   3rd Qu.:0.4000   3rd Qu.:0.3900   3rd Qu.: 8.100
##  Max.   :15.900   Max.   :1.5800   Max.   :1.6600   Max.   :65.800
##    chlorides      free.sulfur.dioxide total.sulfur.dioxide   density
##  Min.   :0.00900  Min.   :  1.00      Min.   :  6.0       Min.   :0.9871
##  1st Qu.:0.03800  1st Qu.: 17.00      1st Qu.: 77.0       1st Qu.:0.9923
##  Median :0.04700  Median : 29.00      Median :118.0       Median :0.9949
##  Mean   :0.05603  Mean   : 30.53      Mean   :115.7       Mean   :0.9947
##  3rd Qu.:0.06500  3rd Qu.: 41.00      3rd Qu.:156.0       3rd Qu.:0.9970
##  Max.   :0.61100  Max.   :289.00      Max.   :440.0       Max.   :1.0390
##        pH           sulphates        alcohol          quality
##  Min.   :2.720   Min.   :0.2200   Min.   : 8.00   Min.   :3.000
##  1st Qu.:3.110   1st Qu.:0.4300   1st Qu.: 9.50   1st Qu.:5.000
##  Median :3.210   Median :0.5100   Median :10.30   Median :6.000
##  Mean   :3.219   Mean   :0.5313   Mean   :10.49   Mean   :5.818
##  3rd Qu.:3.320   3rd Qu.:0.6000   3rd Qu.:11.30   3rd Qu.:6.000
##  Max.   :4.010   Max.   :2.0000   Max.   :14.90   Max.   :9.000
##     color
##  Length:6497
##  Class :character
##  Mode  :character
##
##
##
```

** PCA using prcomp that uses the singular value decomposition (SVD)**

```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     1.7407 1.5792 1.2475 0.98517 0.84845 0.77930 0.72330
## Proportion of Variance 0.2754 0.2267 0.1415 0.08823 0.06544 0.05521 0.04756
## Cumulative Proportion  0.2754 0.5021 0.6436 0.73187 0.79732 0.85253 0.90009
##                           PC8    PC9   PC10    PC11
## Standard deviation     0.70817 0.58054 0.4772 0.18119
## Proportion of Variance 0.04559 0.03064 0.0207 0.00298
## Cumulative Proportion  0.94568 0.97632 0.9970 1.00000
```

** Interpreting the components **

```
##                   PC1   PC2
## fixed.acidity    -0.24  0.34
## volatile.acidity -0.38  0.12
## citric.acid       0.15  0.18
```

```
## residual.sugar         0.35  0.33
## chlorides             -0.29  0.32
## free.sulfur.dioxide    0.43  0.07
## total.sulfur.dioxide   0.49  0.09
## density               -0.04  0.58
## pH                    -0.22 -0.16
## sulphates             -0.29  0.19
## alcohol               -0.11 -0.47
```

Summary of the loadings

```
##                 Properties         PC1         PC2         PC3         PC4
## 1            fixed.acidity -0.23879890  0.33635454 -0.43430130  0.16434621
## 2         volatile.acidity -0.38075750  0.11754972  0.30725942  0.21278489
## 3              citric.acid  0.15238844  0.18329940 -0.59056967 -0.26430031
## 4           residual.sugar  0.34591993  0.32991418  0.16468843  0.16744301
## 5                chlorides -0.29011259  0.31525799  0.01667910 -0.24474386
## 6       free.sulfur.dioxide  0.43091401  0.07193260  0.13422395 -0.35727894
## 7     total.sulfur.dioxide  0.48741806  0.08726628  0.10746230 -0.20842014
## 8                  density -0.04493664  0.58403734  0.17560555  0.07272496
## 9                       pH -0.21868644 -0.15586900  0.45532412 -0.41455110
## 10               sulphates -0.29413517  0.19171577 -0.07004248 -0.64053571
## 11                 alcohol -0.10643712 -0.46505769 -0.26110053 -0.10680270
##           PC5         PC6         PC7          PC8        PC9        PC10
## 1  -0.1474804 -0.20455371 -0.28307944  0.401235645  0.3440567 -0.281267685
## 2   0.1514560 -0.49214307 -0.38915976 -0.087435088 -0.4969327  0.152176731
## 3  -0.1553487  0.22763380 -0.38128504 -0.293412336 -0.4026887  0.234463340
## 4  -0.3533619 -0.23347775  0.21797554 -0.524872935  0.1080032 -0.001372773
## 5   0.6143911  0.16097639 -0.04606816 -0.471516850  0.2964437 -0.196630217
## 6   0.2235323 -0.34005140 -0.29936325  0.207807585  0.3666563  0.480243340
## 7   0.1581336 -0.15127722 -0.13891032  0.128621319 -0.3206955 -0.713663486
## 8  -0.3065613  0.01874307 -0.04675897  0.004831136  0.1128800 -0.003908289
## 9  -0.4533764  0.29657890 -0.41890702 -0.028643277  0.1278367 -0.141310977
## 10 -0.1365769 -0.29692579  0.52534311  0.165818022 -0.2077642  0.045959499
## 11 -0.1888920 -0.51837780 -0.10410343 -0.399233887  0.2518903 -0.205053085
##            PC11
## 1  -0.3346792663
## 2  -0.0847718098
## 3   0.0011089514
## 4  -0.4497650778
## 5  -0.0434375867
## 6   0.0002125351
## 7   0.0626848131
## 8   0.7151620723
## 9  -0.2063605036
## 10 -0.0772024671
## 11  0.3357018784

##                 Properties   PC1
## 1    total.sulfur.dioxide  0.49
## 2     free.sulfur.dioxide  0.43
## 3          residual.sugar  0.35
## 4             citric.acid  0.15
## 5                 density -0.04
## 6                 alcohol -0.11
```

```
## 7                      pH -0.22
## 8          fixed.acidity -0.24
## 9             chlorides -0.29
## 10            sulphates -0.29
## 11      volatile.acidity -0.38
```

** PC1 seems to give more positive loadings to sulfur dioxide and residual sugar while chlorides, sulphates & acidity have negative loadings**

```
##                  Properties   PC2
## 1                  density  0.58
## 2            fixed.acidity  0.34
## 3            residual.sugar  0.33
## 4                chlorides  0.32
## 5                sulphates  0.19
## 6              citric.acid  0.18
## 7         volatile.acidity  0.12
## 8     total.sulfur.dioxide  0.09
## 9      free.sulfur.dioxide  0.07
## 10                      pH -0.16
## 11                  alcohol -0.47
```

PC2 gives more positive loading to density and high negative to alcohol content

** PC space, i.e. the space of summary variables we've created**
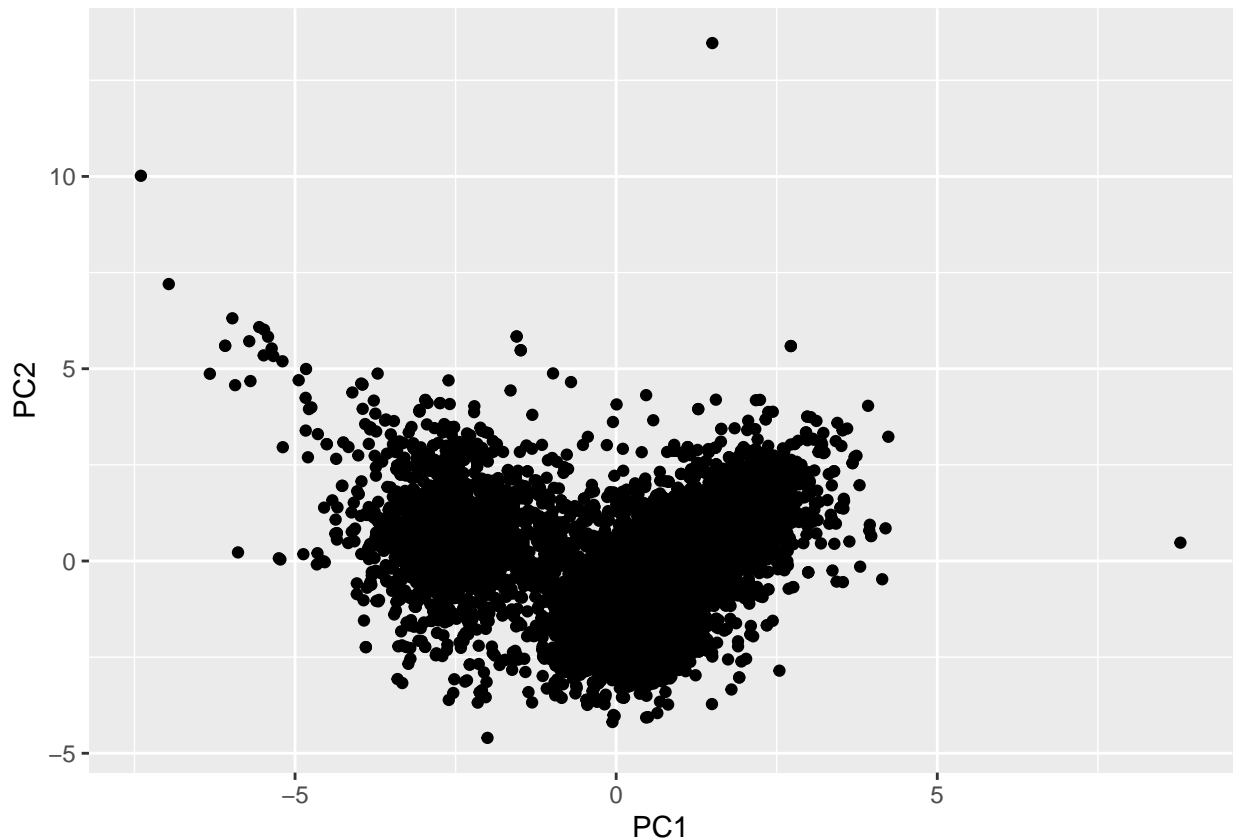

- We can see that there are two clusters emerging. One to the left having more negative PC1 & another to the right having more positive PC1

10

- As we do not have access to the wine color, we cannot say exactly which of these two clusters correspond to red & White wine
- However, the difference in the loadings for each feature by the principal components implies that these two clusters are distinct

Categorize quality

```
##
## High Quality  Low Quality
##          4113          2384
```
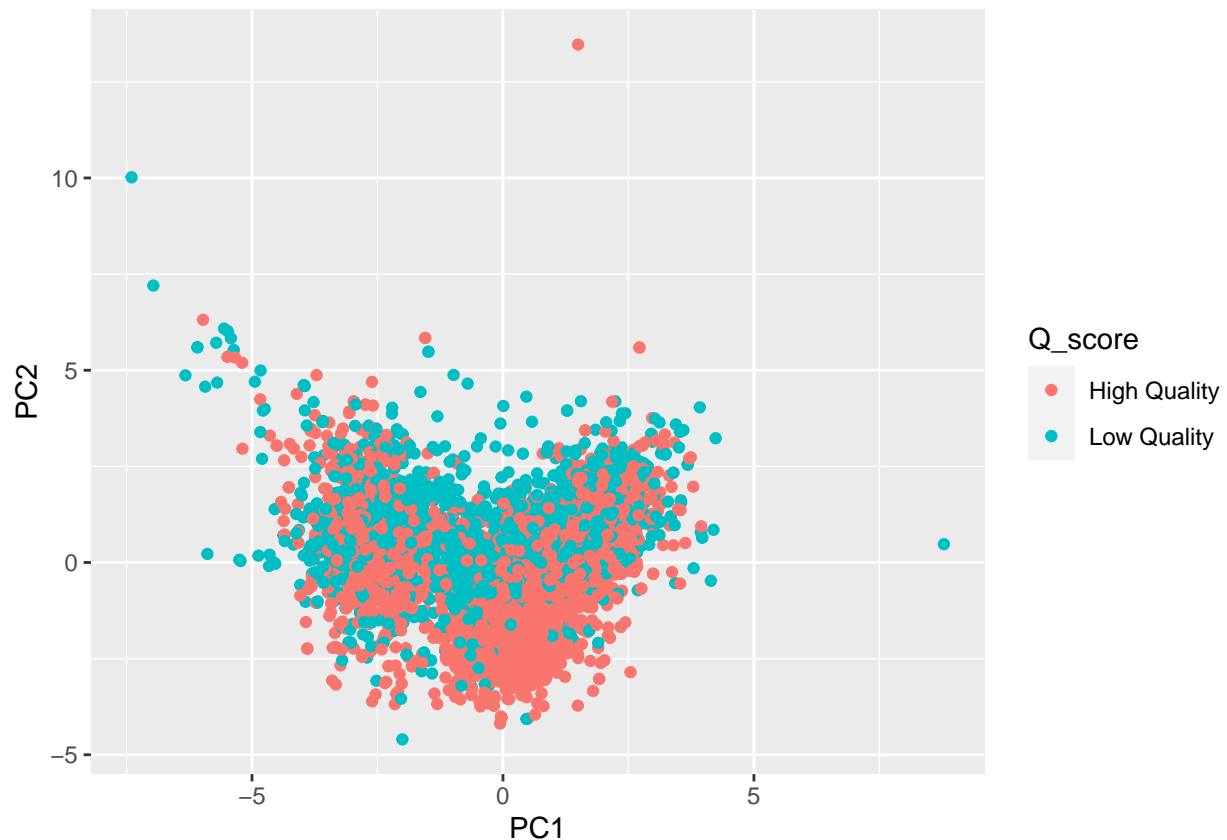

It doesn't look like the principal components can do a good job of separating high & low quality wines.

Clusters

```
##       fixed.acidity     volatile.acidity          citric.acid
##          6.85167903           0.27458385           0.33524928
##       residual.sugar            chlorides  free.sulfur.dioxide
##          6.39402555           0.04510424          35.52152864
## total.sulfur.dioxide             density                   pH
##        138.45848785           0.99400486           3.18762464
##           sulphates              alcohol
##          0.48880511          10.52235888
##
##       fixed.acidity     volatile.acidity          citric.acid
##           8.2895922            0.5319416            0.2695435
##       residual.sugar            chlorides  free.sulfur.dioxide
##           2.6342666            0.0883238           15.7647596
```

```
## total.sulfur.dioxide               density                    pH
##           48.6396835             0.9967404             3.3097200
##            sulphates               alcohol
##            0.6567194            10.4015216

##       volatile.acidity             chlorides             sulphates
##             -1.5631878            -1.2336600            -1.1284118
##           fixed.acidity               density                    pH
##             -1.1091297            -0.9122427            -0.7593601
##               alcohol           citric.acid         residual.sugar
##              0.1013131             0.4521520             0.7902299
##   free.sulfur.dioxide total.sulfur.dioxide
##             1.1130951             1.5890987
```

Table clusters

```
##
##      red white
##   1   24  4830
##   2 1575    68

##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1           7.4             0.70        0.00            1.9     0.076
## 2           7.8             0.88        0.00            2.6     0.098
## 3           7.8             0.76        0.04            2.3     0.092
## 4          11.2             0.28        0.56            1.9     0.075
## 5           7.4             0.70        0.00            1.9     0.076
## 6           7.4             0.66        0.00            1.8     0.075
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  11                   34  0.9978 3.51      0.56     9.4
## 2                  25                   67  0.9968 3.20      0.68     9.8
## 3                  15                   54  0.9970 3.26      0.65     9.8
## 4                  17                   60  0.9980 3.16      0.58     9.8
## 5                  11                   34  0.9978 3.51      0.56     9.4
## 6                  13                   40  0.9978 3.51      0.56     9.4
##   quality color      dist1     dist2
## 1       5   red 107.40131  15.47603
## 2       5   red  72.34191  20.57183
## 3       5   red  87.02246   5.48984
## 4       6   red  80.86079  11.83767
## 5       5   red 107.40131  15.47603
## 6       5   red 101.11543   9.21505
```
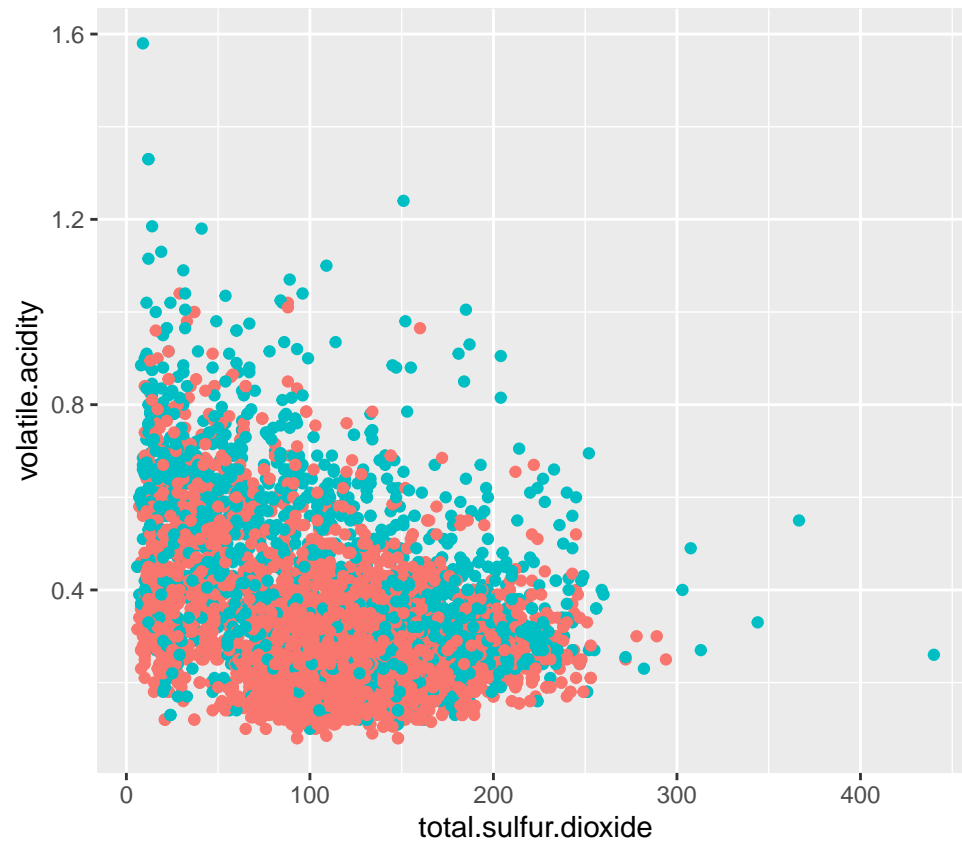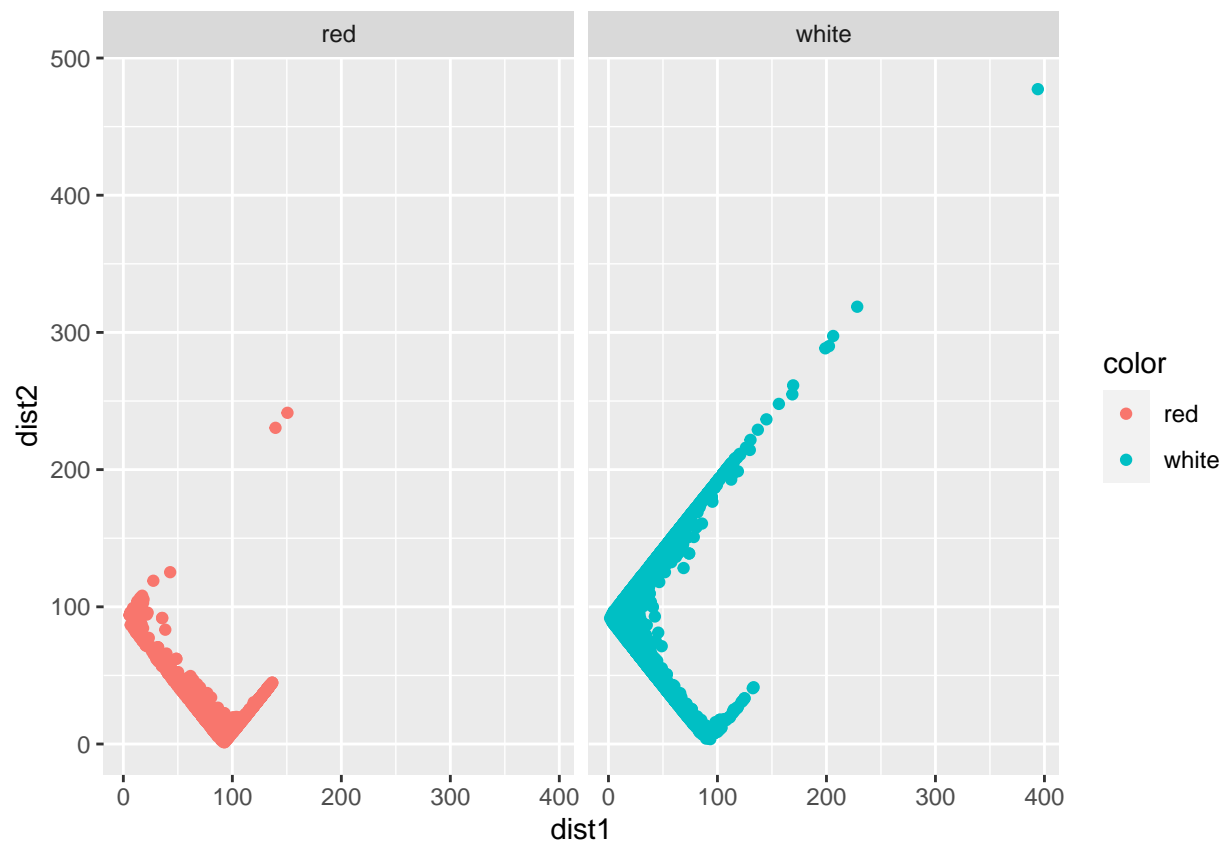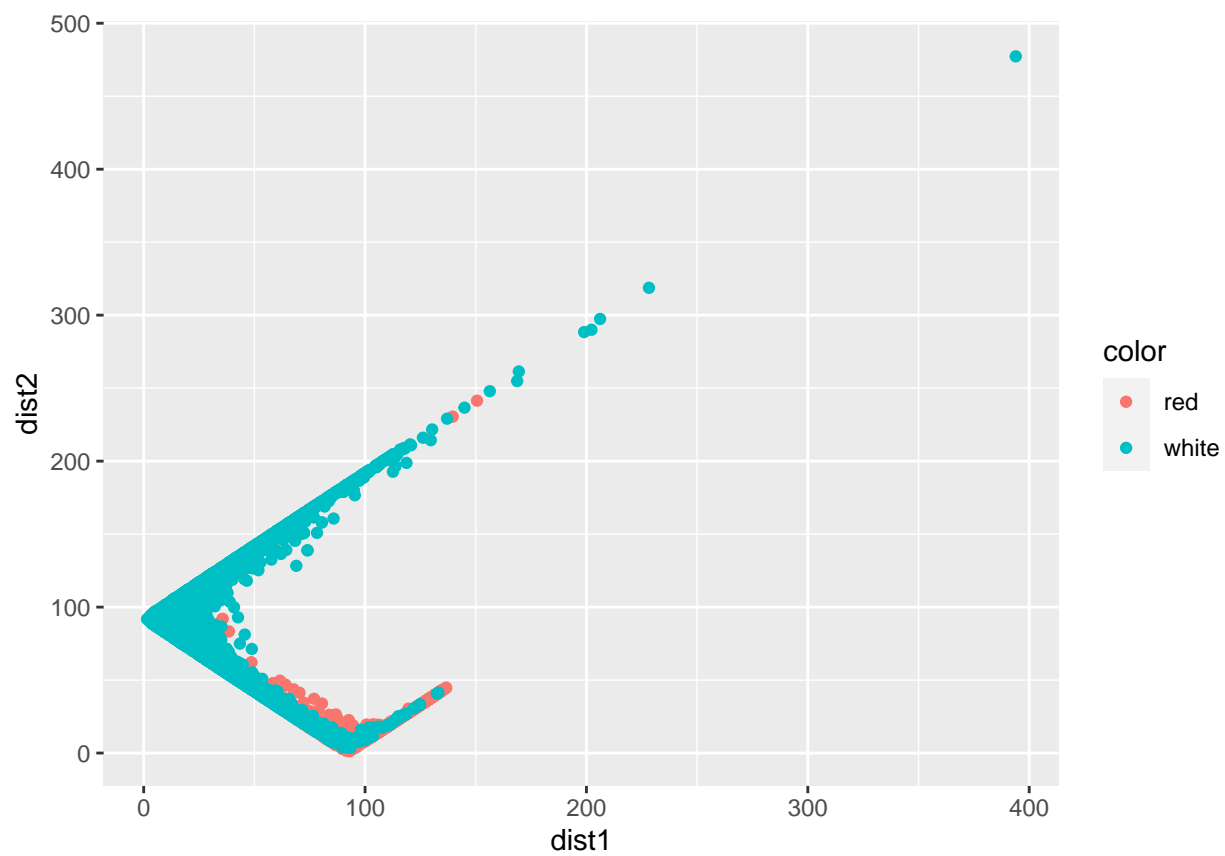
Using the two most contrasting features for the 2 clusters as x & y to visualize
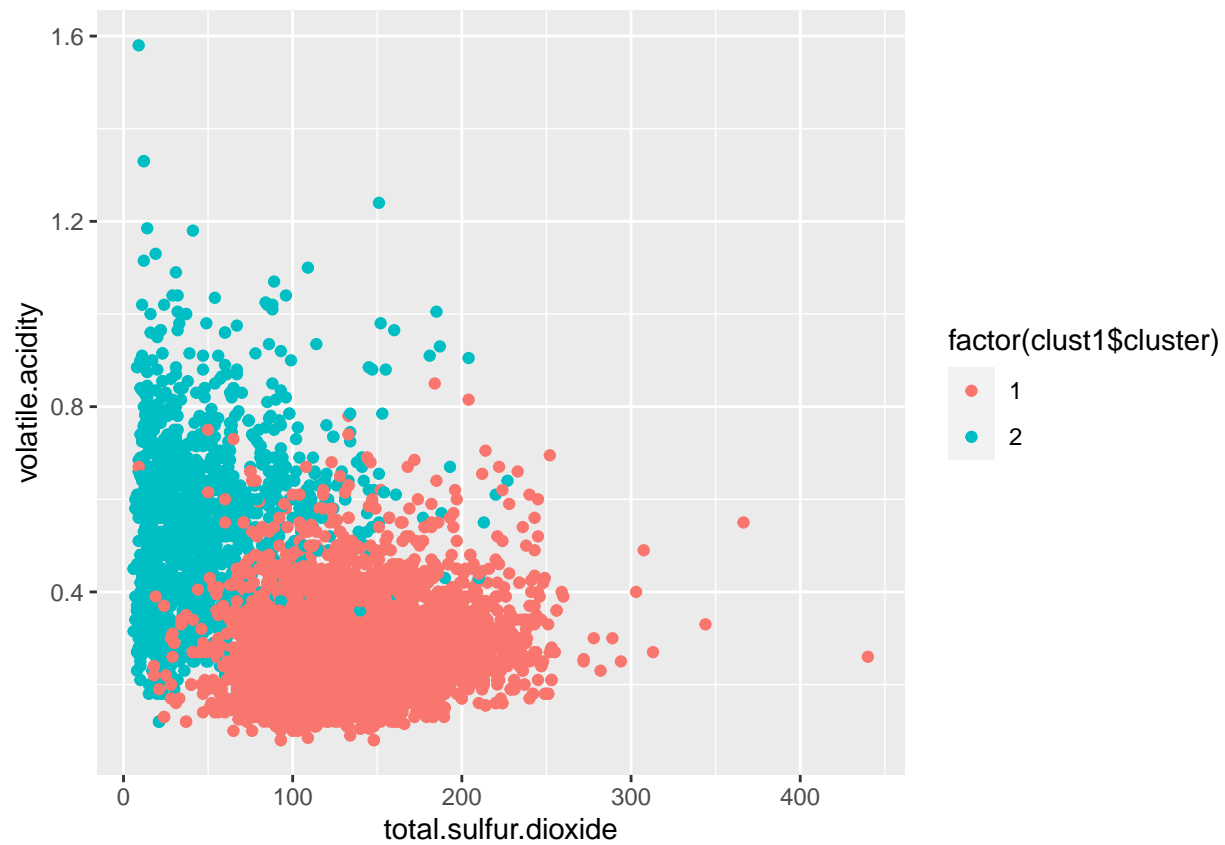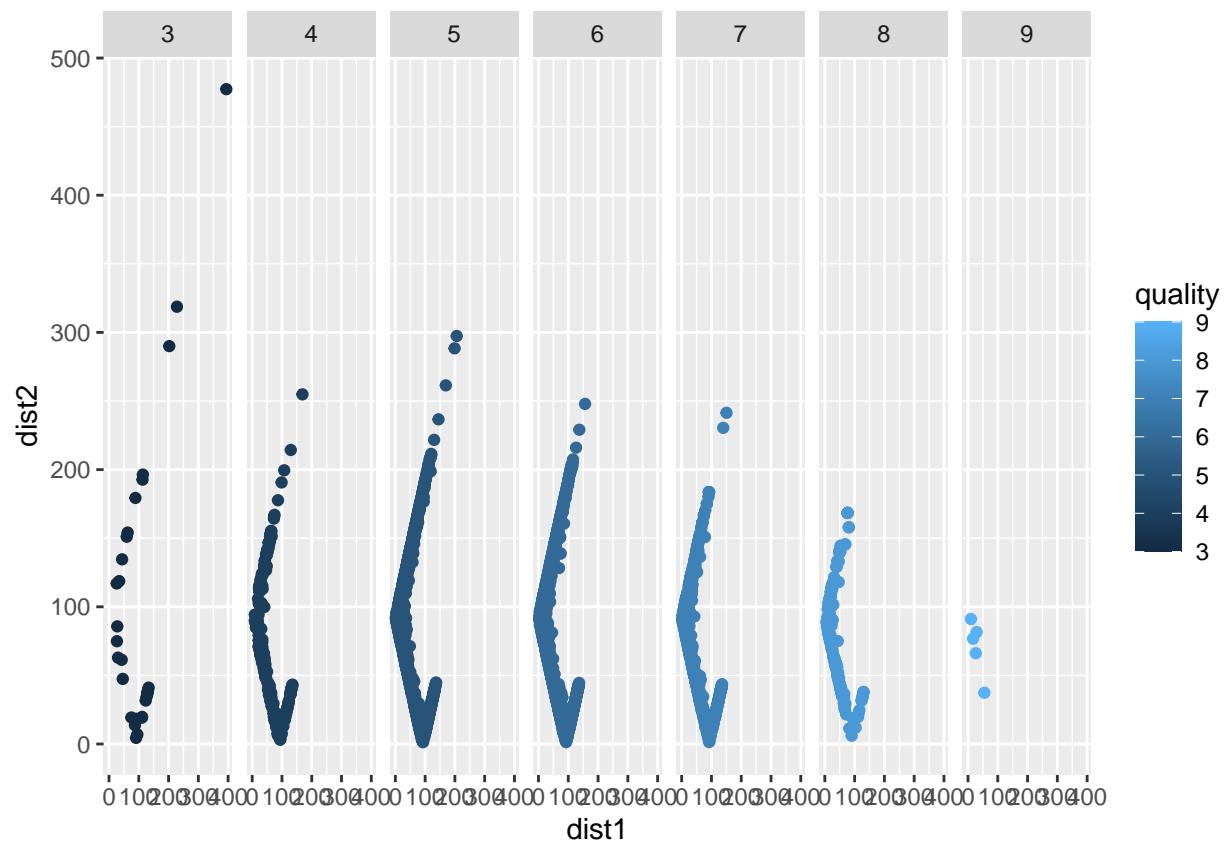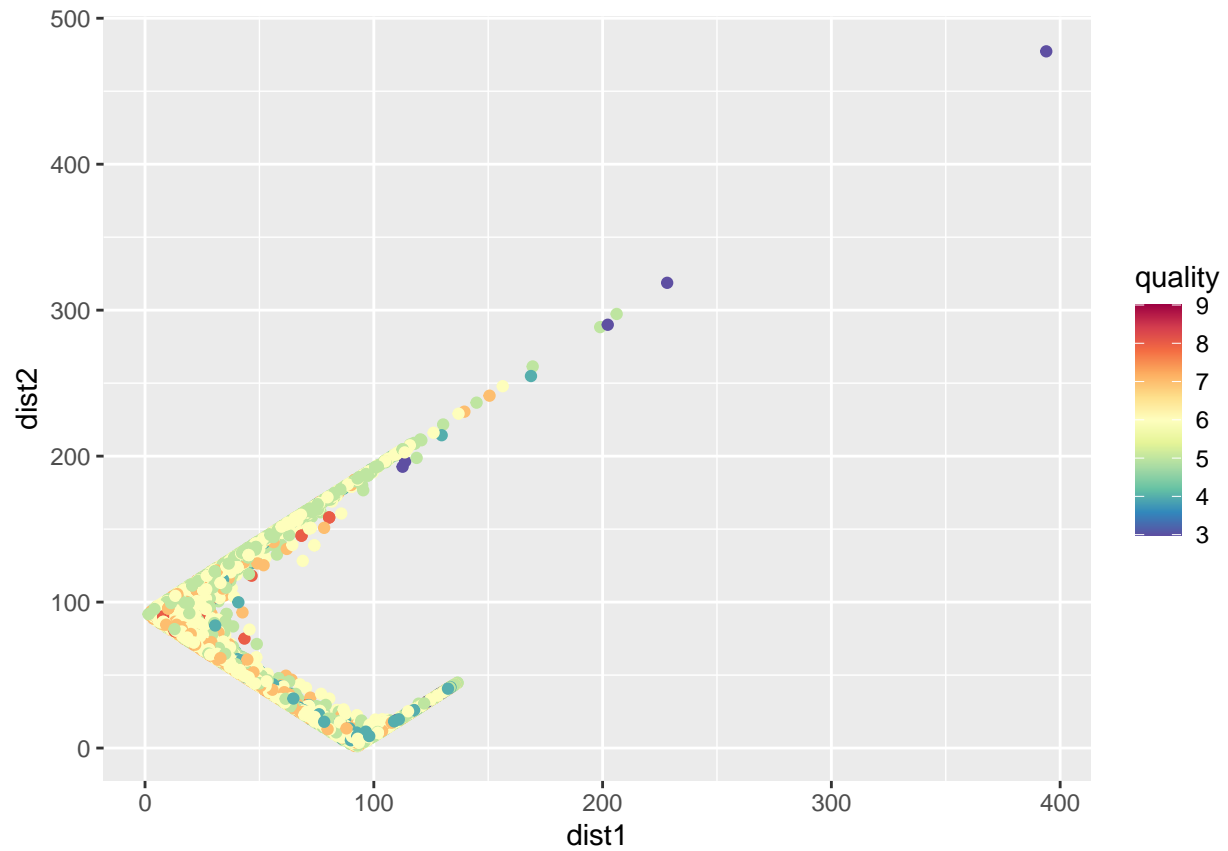
PLot for High and Low Quality Wine

** Summary**

- PCA is the best dimensionality reduction technique makes more sense to me for this data
- PCA Counteracts the issues of high-dimensional data
- PCA improves performance at a very low cost of model accuracy